



A note on the invariance of the distribution of the maximum

Fosgerau, Mogens; Lindberg, Per Olov; Mattsson, Lars Göran; Weibull, Jörgen

Published in:
Journal of Mathematical Economics

DOI:
[10.1016/j.jmateco.2017.10.005](https://doi.org/10.1016/j.jmateco.2017.10.005)

Publication date:
2018

Document version
Peer reviewed version

Document license:
[CC BY-NC-ND](#)

Citation for published version (APA):
Fosgerau, M., Lindberg, P. O., Mattsson, L. G., & Weibull, J. (2018). A note on the invariance of the distribution of the maximum. *Journal of Mathematical Economics*, 74, 56-61. <https://doi.org/10.1016/j.jmateco.2017.10.005>

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/321076552>

A note on the invariance of the distribution of the maximum

Article in *Journal of Mathematical Economics* · November 2017

DOI: 10.1016/j.jmateco.2017.10.005

CITATION

1

READS

114

4 authors:



Mogens Fosgerau

University of Copenhagen

146 PUBLICATIONS 2,358 CITATIONS

[SEE PROFILE](#)



Per Olov Lindberg

KTH Royal Institute of Technology

56 PUBLICATIONS 814 CITATIONS

[SEE PROFILE](#)



Lars-Göran Mattsson

KTH Royal Institute of Technology

67 PUBLICATIONS 2,019 CITATIONS

[SEE PROFILE](#)



Jörgen W. Weibull

Stockholm School of Economics

157 PUBLICATIONS 8,472 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Travel time variability [View project](#)



Discrete choice theory [View project](#)

A note on the invariance of the distribution of the maximum

MOGENS FOSGERAU, PER OLOV LINDBERG, LARS-GÖRAN MATTSSON
AND JÖRGEN WEIBULL*

September 26, 2017

ABSTRACT. Many models in economics involve discrete choices where a decision-maker selects the best alternative from a finite set. Viewing the array of values of the alternatives as a random vector, the decision-maker draws a realization and chooses the alternative with the highest value. The analyst is then interested in the choice probabilities and in the value of the best alternative. The random vector has the *invariance property* if the distribution of the value of a specific alternative, conditional on that alternative being chosen, is the same, regardless of which alternative is considered. This note shows that the invariance property holds if and only if the marginal distributions of the random components are positive powers of each other, even when allowing for quite general statistical dependence among the random components. We illustrate the analytical power of the invariance property by way of examples.

Keywords: Discrete choice, random utility, extreme value, leader-maximum, invariance.

JEL codes: C10, C25, D01.

Many models in economics involve discrete choices where a decision-maker chooses the alternative with the highest value from a finite set of options with random values. Much analytical power can be gained if the maximum value is statistically independent of the choice. Equivalently, as we shall see, we may require that the distribution of the maximum, conditional on alternative i being chosen, does not depend i . We characterize this invariance property while allowing the random values of the alternatives to be statistically dependent.

Our main result, Theorem 1, concerns multivariate distributions that have multivariate extreme-value copulas. For such a distribution, the invariance property holds if and only if the univariate marginal distributions are positive powers of each other.

**Fosgerau*: University of Copenhagen, Denmark. *Lindberg*: Swedish National Road and Transport Research Institute, and Centre for Transport Studies, KTH Royal Institute of Technology, Sweden. *Mattsson*: Department of Transport Science, and Centre for Transport Studies, KTH Royal Institute of Technology. *Weibull*: Department of Economics, Stockholm School of Economics, Department of Mathematics, KTH Royal Institute of Technology, and Institute for Advanced Study in Toulouse. This research has been supported by the European Research Council, and the Knut and Alice Wallenberg Research Foundation.

A special case of the “if” part of this result was given by Resnick and Roy (1990b), who showed that certain multivariate distributions with Gumbel marginals have the invariance property. Mattsson, Weibull and Lindberg (2014) generalized this to the “if” claim in the present Theorem 1. For the independent case, Resnick and Roy (1990a) proved the “only if” claim for the special case of two non-negative random variables.

Researchers in economics and in mathematical psychology have treated the topic of invariance a number of times. However, with few exceptions, earlier research assumes statistical independence between the components of the random vector in question. Strauss (1979, Theorem 4) showed that the multinomial logit model is the only additive random utility model with statistically independent error terms for which the invariance property holds. De Palma and Kilani (2007) revisited the additive random utility case with independent error terms, and proved that for error terms that are i.i.d. with finite expectation, a weaker invariance property, namely invariance in terms of expected values is equivalent to the error terms being Gumbel distributed. In an innovative study, Train and Wilson (2008) used the invariance property of multinomial logit models for maximum-likelihood estimation in a class of combined stated and revealed preference experiments.

The invariance property is by no means evident or always realistic. The “only if” part of our main result shows precisely what subclass of distributions, within the wide class considered here, the researcher then has to avoid if invariance appears unrealistic or unnatural in the application at hand. In particular, the independent multivariate normal distribution does not have the invariance property except in the i.i.d. case (see example illustrated in Figure 1 below), whereas the multivariate extreme-value distribution with Gumbel margins, used in the McFadden class of GEV models, does have the invariance property. However, the invariance property has great technical advantages. It implies specific functional forms and enhances analytical tractability in many applications, such as discrete choice with or without an outside option, rent seeking, patent races, and innovation contests, see Section 3. In discrete choice applications there are often outside options. This makes invariance of the choice probabilities with respect to outside options desirable, since outside options can then (and only then) be ignored. Section 1 provides general definitions and some preliminaries, and our main results are presented in Section 2.

1. DEFINITIONS AND PRELIMINARIES

Let \mathbb{N} be the positive integers, \mathbb{R} the reals and \mathbb{R}_+ the non-negative reals. We write vectors in boldface. A univariate function applied to a vector is understood to mean component-wise application, i.e. $\ln \mathbf{x} = (\ln x_1, \dots, \ln x_n)$. Denote by \mathcal{F} the class of cumulative distribution functions (CDFs) F such that $F : \mathbb{R}^n \rightarrow [0, 1]$, for some integer $n > 1$, is twice continuously differentiable and has positive density f with support $D^n \subseteq \mathbb{R}^n$, where D is a nonempty and (bounded or unbounded) interval. For convenience we will subsequently refer to D^n as “the support”. Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector distributed according to some such F . Let $\hat{X} = \max_i X_i$ be

the maximum of the random vector and let \hat{F} denote its CDF. We write F_i for the partial derivative of F with respect to its i^{th} variable and $F^{(i)}$ for the i^{th} marginal distribution of the multivariate distribution F . Define the selection $\xi \in I = \{1, \dots, n\}$ by $\xi = \arg \max_i X_i$, where the latter set with probability one is a singleton. Let $q_i = P(\xi = i)$. One may also consider the distribution of the maximum conditional on the selection of a particular alternative $i \in I$: $\hat{F}^{(i)}(t) = P(\hat{X} \leq t \mid \xi = i)$. We will denote this conditional random variable \hat{X}_i .¹ In a discrete choice setting, $\hat{F}^{(i)}$ is the distribution of achieved (or experienced) utility, conditional on the choice of alternative $i \in I$, and q_i is its choice probability.

It is relatively straightforward in this setting to prove the following three equalities:²

$$\begin{cases} q_i = \int F_i(s, \dots, s) ds > 0 & \forall i \in I \\ \hat{F}^{(i)}(t) = q_i^{-1} \cdot \int_{-\infty}^t F_i(s, \dots, s) ds & \forall i \in I, t \in \mathbb{R} \\ \hat{F}(t) = F(t, \dots, t) & \forall t \in \mathbb{R}. \end{cases} \quad (1)$$

We note that the quantities in (1) only depend on how the CDF $F \in \mathcal{F}$ behaves near the diagonal of its domain.

By “invariance” we mean that the conditional distributions $\hat{F}^{(i)}$ are identical across alternatives $i \in I$:

Definition 1. A multivariate distribution F has the **Invariance Property** if $\hat{F}^{(i)} = \hat{F}$ for all $i \in I$. A random vector \mathbf{X} has the **Invariance Property** if its distribution has the **Invariance Property**.

The following proposition provides some observations regarding the Invariance Property.

Proposition 1. Consider a random vector \mathbf{X} with CDF $F \in \mathcal{F}$.

- (i) If \mathbf{X} is exchangeable, then \mathbf{X} has the Invariance Property.³
- (ii) If \mathbf{X} has the Invariance Property and $T : \mathbb{R} \rightarrow \mathbb{R}$ is a strictly increasing function, then also $\mathbf{Y} = (T(X_1), \dots, T(X_n))$ has the Invariance Property.
- (iii) \mathbf{X} has the Invariance Property if and only if the selection ξ and the maximum \hat{X} are statistically independent.

The first two observations are immediate. We note that the first observation includes the i.i.d. case. The third follows from

$$P(\hat{X} > x \wedge \xi = i) = P(\hat{X} > x \mid \xi = i) \cdot P(\xi = i) = P(\hat{X}_i > x) \cdot P(\xi = i).$$

¹Technically, let $\Omega_j = \{\omega \in \Omega : X_j(\omega) \geq X_i(\omega) \forall i \in I\}$. Then $\hat{X}_j : \Omega_j \rightarrow \mathbb{R}$ is defined by $\hat{X}_j(\omega) = X_j(\omega)$.

²See e.g. Lindberg, Eriksson and Mattsson (1995, Lemma 1). The general case is more complex, see Lindberg (2012a). Using results from this paper, the conditions on the functions in \mathcal{F} can be relaxed to just twice continuous differentiability, more in line with the conditions on the linear homogenous G below.

³A random vector is *exchangeable* if its distribution is invariant under permutation of its components.

We will express results in terms of so-called copulas. These are functions on the unit cubes in Euclidean spaces, defined as follows (see Nelsen, 2006, for an excellent introduction):

Definition 2. A copula is any function $C : [0, 1]^n \rightarrow [0, 1]$ such that

- (i) $C(\mathbf{x}) = 0$ if $\prod_{i \in I} x_i = 0$,
- (ii) $C(\mathbf{x}) = x_k$ if $\prod_{i \in I \setminus \{k\}} x_i = 1$,
- (iii) If $\mathbf{x}, \mathbf{y} \in [0, 1]^n$ and $\mathbf{x} \leq \mathbf{y}$, then $V_C([\mathbf{x}, \mathbf{y}]) \geq 0$.

Here $[\mathbf{x}, \mathbf{y}]$ denotes the box $\times_{i=1}^n [x_i, y_i] \subseteq [0, 1]^n$, i.e., the Cartesian product of the intervals $[x_i, y_i]$. Moreover, $V_C([\mathbf{x}, \mathbf{y}])$ is the C -volume of this box, defined as the signed sum of the values $C(\mathbf{v})$ at all vertices \mathbf{v} of $[\mathbf{x}, \mathbf{y}]$, where the sign is positive (negative) if $v_i = x_i$ for an even (odd) number of coordinates $i \in I$. Condition (iii) ensures that the copula assigns non-negative probability mass to any box. By construction, copulas are then CDFs on the unit cube that have uniform marginal distributions.

By Sklar's theorem (e.g. Theorem 2.10.9 in Nelsen, 2006), every multivariate distribution $F : \mathbb{R}^n \rightarrow [0, 1]$ can be written in terms of its marginal distributions $F^{(i)}$ and a copula C , so that

$$F(\mathbf{x}) = C(F^{(1)}(x_1), \dots, F^{(n)}(x_n)) \quad \forall \mathbf{x} \in \mathbb{R}^n. \quad (2)$$

The copula associated with any CDF thus captures the statistical dependence structure of the multivariate distribution in question.

In order to state our main result we define the class of copulas that are associated with multivariate extreme-value (MEV) distributions, to be called *MEV copulas*. A *multivariate extreme-value (MEV) distribution* is any multivariate distribution H with non-degenerate margins that can be obtained as the limit of component-wise maxima as follows. Let $(\mathbf{X}_m)_{m \in \mathbb{N}}$ be a sequence of i.i.d. random vectors, with common multivariate CDF F . Let $\mathbf{a}_m, \mathbf{b}_m \in \mathbb{R}^n, m \in \mathbb{N}$ be normalizing vectors such that all components of all vectors \mathbf{a}_m are positive. Using component-wise maximization, multiplication and division to ease on notation, assume the following limit exists and let H be the limiting multivariate CDF:

$$\lim_{m \rightarrow \infty} P\left(\frac{\max_{1 \leq l \leq m} \mathbf{X}_l - \mathbf{b}_m}{\mathbf{a}_m} \leq \mathbf{x}\right) = \lim_{m \rightarrow \infty} [F(\mathbf{a}_m \mathbf{x} + \mathbf{b}_m)]^m = H(\mathbf{x}).$$

Then H is multivariate extreme-value (see Joe, 1997).⁴ This is a natural class of distributions for maximization problems such as when agents make optimal choices, just as the normal distribution is natural for problems involving averages of random variables. As shown in Joe (1997, p. 173-174), an MEV copula C must satisfy

⁴More precisely, if $\mathbf{X}_l = (X_{l1}, \dots, X_{ln})$, then $\frac{\max_{1 \leq l \leq m} \mathbf{X}_l - \mathbf{b}_m}{\mathbf{a}_m}$ is the vector with j^{th} component $\frac{\max_{1 \leq l \leq m} X_{lj} - b_{mj}}{a_{mj}}$, and $\mathbf{a}_m \mathbf{x}$ is the vector with j^{th} component $a_{mj} x_j$.

$C(\mathbf{x}^\alpha) = [C(\mathbf{x})]^\alpha$ for all $\mathbf{x} \in (0, 1)^n$ and $\alpha > 0$. It can be shown that the converse holds, that is, MEV copulas are exactly those that satisfy this homogeneity property. This equivalence follows from

Lemma 1. *A copula C is a copula of an MEV distribution if and only if it is of the form*

$$C(\mathbf{x}) = \exp(-G(-\ln \mathbf{x})) \quad \forall \mathbf{x} \in (0, 1)^n \quad (3)$$

for some function G that is homogeneous of degree 1.

(A proof is given in the Appendix.)

We also note that if $\mathbf{X} = (X_1, \dots, X_n)$ is MEV distributed, then evidently so is any subvector of \mathbf{X} . Hence, the copulas of the subvectors also have the form (3).

Lemma 2. *For $J \subset I$, let $\mathbf{X}_J = (X_j)_{j \in J}$ and similarly for \mathbf{x}_J . Let $G^{(J)}(\mathbf{x}_J) = \lim_{x_i \rightarrow 0 \ \forall i \notin J} G(\mathbf{x})$. If \mathbf{X} is MEV-distributed with CDF of the form (2) with copula C of the form (3), then \mathbf{X}_J has the CDF*

$$F^{(J)}(\mathbf{x}_J) = \exp\left(-G^{(J)}(-\ln F^{(j)}(x_j))_{j \in J}\right).$$

The copula of a vector $\mathbf{X} = (X_1, \dots, X_n)$ of independent random variables is the independence copula $\prod_{i=1}^n x_i$. This is also an MEV copula by Lemma 1 (where G is summation). Our focus on distributions with MEV copulas then includes all independent multivariate distributions.

A simple example of random vectors that have MEV copulas but not the Invariance Property is provided by the multivariate normal distribution with statistically independent variables. Consider a random vector $\mathbf{X} = (X_1, X_2)$ where $X_i \sim N(\mu_i, \sigma^2)$ and X_1 and X_2 are statistically independent. Then F has an MEV copula, the product function, but the Invariance Property holds if and only if $\mu_1 = \mu_2$. This is due to the thin tails of the normal distribution. This failure of invariance is intuitively evident when μ_1 and μ_2 are far apart. Suppose $\mu_1 \ll \mu_2$. Then the probability distribution of \hat{X}_2 is almost the same as that of X_2 , since the probability for $X_1 > X_2$ is small. By contrast, the distribution of \hat{X}_1 is very different from both that of X_1 and X_2 ; its distribution \hat{F}_1 represents an upward biased sample from X_1 . The diagram below illustrates this for $\mu_1 = 0$, $\mu_2 = 2$ and $\sigma^2 = 1$. The thin curves are the densities of X_1 (solid) and X_2 (dashed). The thick curves are the densities for \hat{X}_1 (solid) and \hat{X}_2 (dashed).

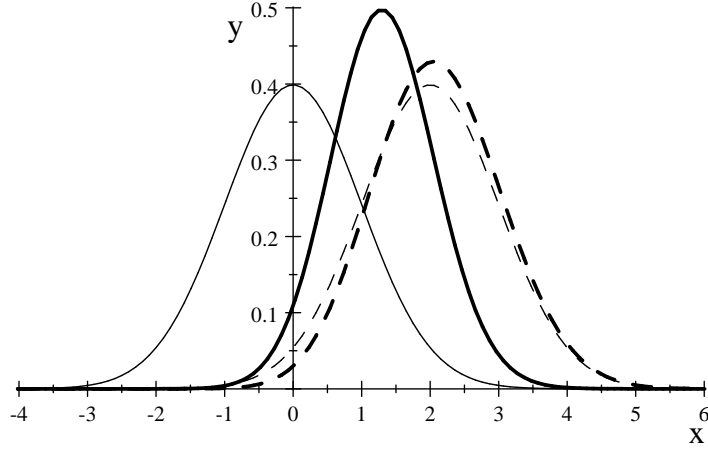


Figure 1: A counter-example to the Invariance Property.

Consider any random vector $\mathbf{X} = (X_1, X_2)$ where X_1 and X_2 are independent, X_1 has CDF $F^{(1)}$ and X_2 has CDF $F^{(2)} = [F^{(1)}]^\alpha$ for some $\alpha > 0$. As follows from Theorem 1 below, \hat{X} , \hat{X}_1 and \hat{X}_2 all have the same distribution, which by (1) is $[F^{(1)}]^{1+\alpha}$. This is then an example of a random vector that has an MEV copula (because the components are independent) and also the Invariance Property. This holds, for example, if one of the components is normal. The diagram shows $F^{(1)}$ (thin), $F^{(2)}$ (thin dashed), and \hat{F} (solid), when X_1 is standard normal and $\alpha = 2$. This is then an example of how the Invariance Property can be achieved based on the normal distribution and an MEV copula.

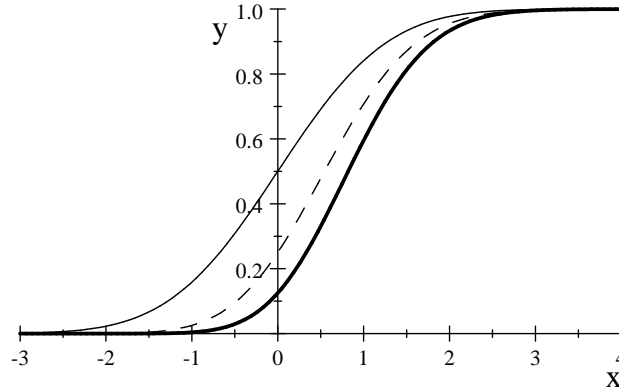


Figure 2: An example of powers of the normal distribution, with the Invariance Property.

2. RESULTS

We are now in a position to state the main result of this note.

Theorem 1. *Consider any $F \in \mathcal{F}$ satisfying (2) for an MEV copula C that is twice continuously differentiable with positive first-order partial derivatives. Then F and*

all the multivariate marginal distributions of F have the Invariance Property if and only if for each $i \in I$ there exists an $\alpha_i > 0$ such that $F^{(i)} = [F^{(1)}]^{\alpha_i}$.

(A proof is given in the Appendix.)

The assumption in the Theorem that C is twice continuously differentiable is a bit stronger than needed, but simplifies the statement and proof. We also note that it follows from Lemma 2 that all the multivariate marginal distributions of a CDF F (as in the Theorem) have the Invariance Property if F has it.

Examples of marginal distributions $F^{(i)}$ that satisfy the power equation in the Theorem are Gumbel distributions with the same standard deviation. The same holds for Fréchet distributions with the same shape parameter. Indeed, Gumbel distributions are the basis of the much used multinomial logit model, and its generalization to GEV choice models, a literature pioneered by McFadden (1974, 1978, 1981). For examples of choice models based on Fréchet distributions, see e.g. Eaton and Kortum (2002).

Similarly, consider minimization instead of maximization, and a multivariate survival function that satisfies the corresponding hypothesis of Theorem 1. If it has the corresponding Invariance Property, and if one of the marginal distributions is Weibull or Pareto, then the other marginal distributions are so too.

We also note that, except for some special cases, the Invariance Property is not preserved when one mixes distributions that are invariant. For example, a convex combination of different Gumbel distributions does not have the Invariance Property.⁵ Such distributions occur in mixed multinomial logit models (see e.g. McFadden and Train, 2000).

Theorem 1 characterizes the Invariance Property for those distributions in \mathcal{F} that have twice continuously differentiable MEV copulas with positive first-order partial derivatives. Could it be that invariance of a probability distribution is *equivalent* to the distribution having an MEV copula and marginal distributions that are powers of each other? That conjecture is false. This follows from the next proposition, according to which one may change any multivariate distribution at points outside a neighborhood of the diagonal of its domain without affecting its Invariance Property. Let L be the diagonal of D^n ; $L = \{\mathbf{x} \in D^n : x_1 = \dots = x_n\}$.

Proposition 2. *Consider any distributions $F, \tilde{F} \in \mathcal{F}$ with the same support D^n . If F has the Invariance Property and $\tilde{F} = F$ on an open neighborhood of the diagonal L , then also \tilde{F} has the Invariance Property.*

To prove this proposition, it is sufficient to note that by (1), the CDFs of the unconditional and conditional maxima, $\hat{F}, \hat{F}^{(1)}, \dots, \hat{F}^{(n)}$, depend only on the values of the joint distribution F and its partial derivatives on the diagonal. The CDFs of the unconditional and conditional maxima corresponding to \tilde{F} are then also $\hat{\tilde{F}}, \hat{\tilde{F}}^{(1)}, \dots, \hat{\tilde{F}}^{(n)}$.

The next proposition establishes that the Invariance Property is preserved under aggregation of components to blocks represented by their maximal member.

⁵Except perhaps in very special cases

Proposition 3. *Suppose that $\mathbf{X} = (X_1, \dots, X_n)$ has the Invariance Property, and consider any partition of I into k subsets. For each subset I_t let $Y_t = \max_{i \in I_t} (X_i)$. Then $\mathbf{Y} = (Y_1, \dots, Y_k)$ has the Invariance Property.*

(A proof is given in the Appendix.)

Considering results related to our Theorem 1, Strauss (1979) in his Theorem 5 (without proof⁶) and Lindberg, Eriksson and Mattsson (1995) in their Theorem 4 (with proof) claim that if the components of a random vector \mathbf{X} are of the additive form $X_i = \nu_i + Z_i$, where the ν_i 's are scalars and \mathbf{Z} is a random vector with CDF F , then the Invariance Property holds for all vectors $\boldsymbol{\nu}$ if and only if $F(\mathbf{z}) = \phi[G(e^{-\mathbf{z}})]$ for some positive homogenous function G (with positive degree) and some function $\phi : \mathbb{R}_+ \rightarrow [0, 1]$ such that F is a CDF. All three theorems imply that GEV choice models have the Invariance Property. However, only our Theorem 1 is applicable to, e.g., multivariate distributions with Fréchet marginals or, in the case of minimization, to distributions with Weibull or Pareto marginals. As for the only-if-part, their theorems require the Invariance Property to hold for all $\boldsymbol{\nu} \in \mathbb{R}^n$, whereas our Theorem 1 has no such requirement. Also Marley (1989) derived the Invariance Property within what he termed a “horse race” model, a model that can be viewed as a transformed version of the above model of Strauss.⁷

3. APPLICATIONS

This section discusses two examples in which the Invariance Property is useful. Throughout we rely on the equivalence of the Invariance Property to the independence of the selection ξ and the maximum \hat{X} .

3.1. Discrete choice with outside option. The probabilistic basis of a discrete choice model is a random vector \mathbf{X} of (indirect) utilities, each associated with an alternative i in some set I . Suppose there is also an outside option (say, not to buy from any one of the n sellers in the market). The decision-maker, who we may think of as a consumer, chooses the alternative with the highest utility, including the outside option as a possibility. Such a discrete choice model is, for example, used in models of monopolistic competition, see e.g. Sattinger, 1984, or Perloff and Salop, 1985.

The utility X_0 associated with the outside option acts as a threshold such that the consumer only chooses one of the alternatives in I if the utility of that alternative exceeds that of the outside option. Such situations are easily treated when the random utilities have the Invariance Property.

Suppose that \mathbf{X} has the Invariance Property, and that alternative i is chosen if $X_i = \hat{X}$ and $\hat{X} > X_0$, where $\hat{X} = \max_{i \in I} X_i$ and X_0 is the random utility of the outside option. We henceforth take X_0 to be statistically independent from \mathbf{X} . We still use ξ to denote the alternative in $I = \{1, \dots, n\}$ with maximum utility and let q_i

⁶The proof claimed in Robertson and Strauss (1981) is, unfortunately, erroneous.

⁷See Appendix 2 in Lindberg (2012b).

denote the probability that $\xi = i$. Let $\eta \in \{0\} \cup I$ denote the chosen option among all options. With F_0 being the CDF of X_0 , the probability for the outside option is

$$P(\eta = 0) = P(\hat{X} \leq X_0) = \int F(s, \dots, s) dF_0(s);$$

this probability can be easily calculated from the primitives of the model. Moreover, it follows from Proposition 1 (iii) that, for any $i > 0$,

$$\begin{aligned} P(\eta = i) &= P(\xi = i \wedge \hat{X} > X_0) = \int P(\xi = i \mid \hat{X} > s) \cdot P(\hat{X} > s) dF_0(s) \\ &= q_i \cdot [1 - P(\eta = 0)]. \end{aligned}$$

Hence, the presence of an outside option does not affect the conditional probabilities of the inside options, once the outside option has been rejected by the consumer or decision maker. This holds for any CDF F_0 of the outside option X_0 , as long as the latter is statistically independent of X . In particular, the outside option may be discrete or deterministic.⁸

We now establish the converse: that if the presence of an outside option does not affect the inside choice probabilities, then the inside utilities must have the Invariance Property. Suppose then that value of the outside option is deterministic and that the inside choice probabilities do not depend on the value of the outside option, i.e. $P(\xi = i \mid \hat{X} > x) = P(\xi = i)$ for all $i \in I$ and $x < \sup D$. Then

$$\begin{aligned} P(\xi = i \wedge \hat{X} > x) &= P(\xi = i \mid \hat{X} > x) \cdot P(\hat{X} > x) \\ &= P(\xi = i) \cdot P(\hat{X} > x). \end{aligned}$$

Hence, ξ and \hat{X} are then statistically independent, and thus

$$P(\hat{X} \leq x \mid \xi = i) = P(\hat{X} \leq x)$$

for all $i \in I$ and $x \in \mathbb{R}$. That is, the Invariance Property holds: $\hat{F}_i = \hat{F}$ for all $i \in I$.

3.2. Rent-seeking, innovation contests and patent races. Baye and Hoppe (2003) establish the strategic equivalence between wide classes of rent-seeking games, innovation contests and patent-races. In their innovation-contest game, n firms compete by employing a finite and positive number of scientists, where each scientist costs

⁸Outside options in random utility models have been considered before. To the best of our knowledge, the outside option has then either been (a) deterministic, as in Besanko, Perry and Spady (1990), (b) independently and identically Gumbel-distributed up to location shifts, as in Anderson, de Palma and Thisse (1992), or (c) allowed to be statistically dependent with the inside options in the form of a nested logit model, as in Mattsson and Weibull (1981).

$c > 0$ and independently produces an innovation of random value in $[0, 1]$ with CDF H . All firms pay the costs of their scientists, and the firm with the best idea among all firms wins the value of its best idea. The other firms win nothing.

Let X_i be the value of firm i 's best idea and let $\mathbf{X} = (X_1, \dots, X_n)$. Then \mathbf{X} has the joint CDF $F(\mathbf{x}) = \prod_{i=1}^n F^{(i)}(x_i)$, where $F^{(i)}(x_i) = [H(x_i)]^{s_i}$ and $s_i > 0$ is the number of scientists in firm i . Clearly, $\hat{F} = H^{s_1 + \dots + s_n}$. Moreover, F has an MEV copula (by Lemma 1) that satisfies the hypothesis of Theorem 1 and thus has the Invariance Property. By Proposition 1 (iii), the identity $\xi \in I$ of the firm that wins the contest and the value \hat{X} of the best idea are statistically independent. Hence, the expected profit to any firm i can be expressed as

$$\begin{aligned} \pi_i(\mathbf{s}) &= P(\xi = i) \cdot E(\hat{X}) - cs_i \\ &= \frac{s_i}{s_1 + \dots + s_n} \cdot \left(1 - \int_0^1 [H(t)]^{s_1 + \dots + s_n} dt\right) - cs_i \end{aligned}$$

Treating the R&D inputs s_i as positive real numbers, this representation of profit functions permit game-theoretic analyses of firms' equilibrium investment in R&D inputs. Moreover, this can be generalized to include "synergy" or "critical mass" effects with respect to the number of scientists in a firm, by letting the CDF $F^{(i)}$ of the value of the best idea within any firm i depend in a more general way on the number of scientists in the firm. Hence, let $F^{(i)}$ depend on s_i ; $F^{(i)}(x_i) = K_i(x_i, s_i)$. Such a function K_i could represent the arguably realistic phenomenon that the marginal return, in terms of innovations or discoveries, from an additional scientist may be highest at some intermediate size of a research unit. We could also allow for statistical dependence between distinct firms' best ideas by letting the joint CDF $F(\mathbf{x}) = C(F^{(1)}(x_1), \dots, F^{(n)}(x_n))$ for some MEV copula C that satisfies the differentiability hypothesis of Theorem 1.

To retain statistical independence between the selection of the winning firm and the value of the best idea, for any subset of firms taking part in the competition, we can apply the "only if" part of Theorem 1. More specifically, if such a formulation is to hold for any number of firms and any number of scientists in each firm, the functions K_i need to be of the power form $K_i(x_i, s_i) = [A(x_i)]^{r_i(s_i)}$ for some positive functions r_i and some CDF A .⁹

4. APPENDIX

Proof of Lemma 1. Necessity: Note with Joe (1997, p. 173-174) that an MEV copula C must satisfy $C(\mathbf{x}^\alpha) = [C(\mathbf{x})]^\alpha$ for all $\mathbf{x} \in (0, 1)^n$ and $\alpha > 0$. Then $G(\mathbf{y}) \equiv -\ln C(e^{-\mathbf{y}})$ is homogenous of degree 1. Sufficiency: Suppose C has the form (3). Let F be a CDF with copula C and marginals that are unit exponential on the negative half-axis, i.e. $F_i(x) = e^x$ on $(-\infty, 0]$. Then $F(\mathbf{x}) = C(e^{\mathbf{x}})$. Let $\mathbf{X}_i = (X_{i1}, \dots, X_{in})$

⁹We note that since firms usually can split their R&D units if they wish, the functions r_i can be taken to be super-additive; $r_i(s+t) \geq r_i(s) + r_i(t)$.

be i.i.d. with CDF F , and let $M_j^k = \max_{i \leq k} X_{ij}$. Let $\mathbf{Y}^k = (kM_1^k, \dots, kM_n^k)$, then \mathbf{Y}^k has the CDF

$$\begin{aligned} P(\mathbf{Y}^k \leq \mathbf{x}) &= P[(kM_1^k, \dots, kM_n^k) \leq (x_1, \dots, x_n)] \\ &= P\left[\left(\max_{i \leq k} X_{i1}, \dots, \max_{i \leq k} X_{in}\right) \leq \left(\frac{x_1}{k}, \dots, \frac{x_n}{k}\right)\right] \\ &= \prod_{i \leq k} P\left[X_{i1}, \dots, X_{in} \leq \left(\frac{x_1}{k}, \dots, \frac{x_n}{k}\right)\right] \\ &= [C(e^{x_1/k}, \dots, e^{x_n/k})]^k = C(e^{\mathbf{x}}). \end{aligned}$$

Thus all \mathbf{Y}^k have the same distribution, and hence they converge in distribution to \mathbf{Y}^1 , say, with the same CDF $C(e^{\mathbf{x}})$. Therefore $C(e^{\mathbf{x}})$ is the CDF of an MEV distribution and C is an MEV copula. ■

Proof of Theorem 1. Sufficiency: Theorem 1 in Mattsson, Weibull and Lindberg (2014) states that a CDF $F \in \mathcal{F}$ of the form

$$F(x_1, \dots, x_n) = \exp(-G(-\alpha_1 H(x_1), \dots, -\alpha_n H(x_n))), \quad (4)$$

where G is a homogenous function of degree 1, and H is a univariate CDF, has the Invariance Property. Sufficiency then follows from the observation in Lemma 2 that the multivariate marginal distributions of MEV copulas are also MEV copulas, and that twice differentiability with positive partials is inherited.

Necessity: Consider the CDF

$$P(X_1 \leq x_1, X_2 \leq x_2) = F(x_1, x_2, \infty, \dots, \infty) = C(F^{(1)}(x_1), F^{(2)}(x_2), 1, \dots, 1).$$

By assumption, this CDF has the Invariance Property. As noted in Lemma 2, the copula for (X_1, X_2) inherits the MEV property from C . We may thus ignore the last but two dimensions of F and C , and assume that $|I| = 2$ at no loss of generality, and will henceforth write $\mathbf{X} = (X_1, X_2)$. It remains to show that $F^{(2)} = [F^{(1)}]^\alpha$ for some $\alpha > 0$.

First, by Lemma 1, we may write

$$F(\mathbf{x}) = \exp(-G(-\ln F^{(1)}(x_1), -\ln F^{(2)}(x_2))), \quad (5)$$

where G is homogenous of degree 1 and satisfies the properties necessary for F to be a CDF. Since C by hypothesis is twice continuously differentiable with positive partials, G inherits these properties (except at the origin). To see this, note that $G(y) \equiv -\ln C(e^{-y_1}, e^{-y_2})$.

Second, by Proposition 1, the Invariance Property remains under any strictly increasing transformation of the components. Hence, it is no loss of generality to apply

such a transformation so that $F^{(1)}$ is a Gumbel distribution, $F^{(1)}(x_1) = \exp(-e^{-x_1})$ on \mathbb{R} . By (1) we then have, for all $t \in \mathbb{R}$,

$$\begin{aligned}\hat{F}^{(1)}(t) &= \frac{1}{q_1} \int_{-\infty}^t F(s, s) G_1(e^{-s}, -\ln F^{(2)}(s)) e^{-s} ds = \\ \hat{F}^{(2)}(t) &= \frac{1}{q_2} \int_{-\infty}^t F(s, s) G_2(e^{-s}, -\ln F^{(2)}(s)) \frac{f^{(2)}(s)}{F^{(2)}(s)} ds,\end{aligned}$$

where $f^{(2)} > 0$ is the density of $F^{(2)}$ and q_1 and q_2 are the associated choice probabilities. Differentiation with respect to t gives

$$\frac{1}{q_1} F(t, t) G_1(e^{-t}, -\ln F^{(2)}(t)) \cdot e^{-t} = \frac{1}{q_2} F(t, t) G_2(e^{-t}, -\ln F^{(2)}(t)) \cdot \frac{f^{(2)}(t)}{F^{(2)}(t)}.$$

or (since $F(t, t) > 0$ and $G_i(\mathbf{x}) > 0$ for all positive $\mathbf{x} \in \mathbb{R}^2$):

$$f^{(2)}(t) = \frac{q_2 e^{-t} G_1(e^{-t}, -\ln F^{(2)}(t))}{q_1 G_2(e^{-t}, -\ln F^{(2)}(t))} \cdot F^{(2)}(t) \quad \forall t \in \mathbb{R}. \quad (6)$$

Since F with $F^{(1)}(x_1) \equiv \exp(-e^{-x_1})$ and $F^{(2)} = [F^{(1)}]^\alpha$ for any $\alpha > 0$ has the Invariance Property (by the established sufficiency claim of this theorem), $F^{(2)}$ satisfies this equation, where q_1 and $q_2 = 1 - q_1$ are the associated choice probabilities. Suppose that $\bar{F}(x_1, x_2) = C(F^{(1)}(x_1), \bar{F}^{(2)}(x_2))$ is another invariant CDF, but with $\bar{F}^{(2)} \neq [F^{(1)}]^\alpha$ for all $\alpha > 0$. Then $\bar{F}^{(2)}$ is a solution to (6) for $q_i = \bar{q}_i$, where \bar{q}_i are the choice probabilities associated with \bar{F} . We will show that no such solution $\bar{F}^{(2)}$ exists. For this purpose we first show that, for any $\bar{q}_1 \in (0, 1)$ there exists some $\alpha > 0$ such that $\bar{q}_1 = q_1(\alpha)$, where $q_1(\alpha)$ and $q_2(\alpha) = 1 - q_1(\alpha)$ are the choice probabilities under the CDF F with $F^{(1)}(x_1) \equiv \exp(-e^{-x_1})$ and $F^{(2)} = [F^{(1)}]^\alpha$. This can be established as follows. By Theorem 1 in McFadden (1978),

$$q_1(\alpha) = \frac{G_1(1, \alpha)}{G(1, \alpha)} \quad \text{and} \quad q_2(\alpha) = \frac{\alpha G_2(1, \alpha)}{G(1, \alpha)}$$

(set $V_1 = 0$ and $V_2 = \ln \alpha$ in McFadden's equation (12)). By homogeneity of each G_i one obtains

$$\frac{1 - q_1(\alpha)}{q_1(\alpha)} = \alpha \cdot \frac{G_2(1/(1 + \alpha), \alpha/(1 + \alpha))}{G_1(1/(1 + \alpha), \alpha/(1 + \alpha))}$$

Moreover, since each G_i is continuous and positive on $\Delta = \{\mathbf{y} \in \mathbb{R}_+^2 : y_1 + y_2 = 1\}$ also the ratio G_2/G_1 is continuous and positive on Δ . Since Δ is compact, there exists $a, b > 0$ such that $a \leq G_2(\mathbf{y})/G_1(\mathbf{y}) \leq b$ for all $\mathbf{y} \in \Delta$. Thus, for any $\bar{q}_1 \in (0, 1)$ there exist an $\bar{\alpha} > 0$ such that $q_1(\bar{\alpha}) = \bar{q}_1$.

According to the Picard-Lindelöf Theorem (see, e.g., Theorem 3.1 in Hale, 1969), an ordinary differential equation such as (6) has a unique (local) solution through any given point $(t_0, x_0) \in \mathbb{R} \times (0, 1)$, for $x_0 = F^{(2)}(t_0)$ if

$$\Psi(t, x) = \frac{q_2 e^{-t} G_1(e^{-t}, -\ln x)}{q_1 G_2(e^{-t}, -\ln x)} \cdot x$$

defines a continuous function on $\mathbb{R} \times (0, 1)$ that is locally Lipschitz-continuous in x . Our function Ψ is continuously differentiable in x and thus also locally Lipschitz continuous in x . By uniqueness, the solution $F^{(2)} = [F^{(1)}]^\alpha$ and the hypothesized alternative solution, $\bar{F}^{(2)}$ cannot intersect at any point (where their values are in $(0, 1)$). Hence, $\bar{F}^{(2)}(t)$ either lies above or below our solution for all $t \in \mathbb{R}$. Assume it always lies above: $F^{(2)}(x_2) < \bar{F}^{(2)}(x_2)$ for all x_2 where $F^{(2)}(x_2) < 1$ and $\bar{F}^{(2)}(x_2) > 0$ (the opposite case can be treated in the same way). The random vector \mathbf{X} has the CDF $C(F^{(1)}(x_1), F^{(2)}(x_2))$, where C is the above copula (restricted to $n = 2$). Let $\bar{\mathbf{X}}$ have the CDF $C(F^{(1)}(x_1), \bar{F}^{(2)}(x_2))$. Further, let $\mathbf{Y} = (Y_1, Y_2)$ have the CDF C , and let $F^{-(i)}$ be the inverse of $F^{(i)}$, for $i = 1, 2$ and $\bar{F}^{-(2)}$ that of $\bar{F}^{(2)}$. Then $X_i = F^{-(i)}(Y_i)$ and $\bar{X}_2 = \bar{F}^{-(2)}(Y_2)$. Clearly $\bar{F}^{-(2)} < F^{-(2)}$ on $(0, 1)$. Since $F^{(2)} < \bar{F}^{(2)}$ and \mathbf{Y} has positive density everywhere,

$$\begin{aligned} q_2(\bar{\alpha}) &= P(X_2 \geq X_1) = P[F^{-(2)}(Y_2) \geq F^{-(1)}(Y_1)] \\ &> P[\bar{F}^{-(2)}(Y_2) \geq F^{-(1)}(Y_1)] = P(\bar{X}_2 \geq X_1) = \bar{q}_2, \end{aligned}$$

a contradiction. ■

Proof of Proposition 3. Let \mathbf{X} have the Invariance Property, i.e., $\hat{F}^{(i)} = \hat{F}$ for all i . It is sufficient to establish the proposition for the case when two alternatives, say 1 and 2, are merged via the maximum operation. Define $Y = X_1 \vee X_2$. It is sufficient to show that $P(Y \leq t \mid \xi \in \{1, 2\}) = \hat{F}(t)$. But

$$\begin{aligned} P(Y \leq t \mid \xi \in \{1, 2\}) &= \frac{P([Y \leq t] \wedge [\xi \in \{1, 2\}])}{P(\xi \in \{1, 2\})} \\ &= \frac{P([Y \leq t] \wedge [\xi = 1]) + P([Y \leq t] \wedge [\xi = 2])}{q_1 + q_2} \\ &= \frac{P(Y \leq t \mid \xi = 1) P(\xi = 1) + P(Y \leq t \mid \xi = 2) P(\xi = 2)}{q_1 + q_2} \\ &= \frac{q_1}{q_1 + q_2} \hat{F}^1(t) + \frac{q_2}{q_1 + q_2} \hat{F}^2(t) = \hat{F}(t). \quad \blacksquare \end{aligned}$$

REFERENCES

- [1] Andersson S., A. de Palma, and J. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. MIT Press.
- [2] Baye, M.R. and H.C. Hoppe (2003): “The strategic equivalence of rent-seeking, innovation, and patent-race games”, *Games and Economic Behavior* 44 (2), 217-226.
- [3] Besanko, D., M. Perry, and R. Spady (1990): “The logit model of monopolistic competition: Brand diversity”, *The Journal of Industrial Economics* 38, 397-415.

- [4] de Palma, A., and K. Kilani (2007): “Invariance of conditional maximum utility”, *Journal of Economic Theory* 132 (1), 137-146.
- [5] Eaton, J., and S. Kortum (2002): “Technology, geography, and trade”, *Econometrica* 5, 1741-1779.
- [6] Hale, J.K. (1969): *Ordinary Differential Equations*. New York: Wiley and Sons.
- [7] Joe, H. (1997): *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.
- [8] Lindberg, P.O. (2012a): “Choice probabilities in random utility models – Necessary and sufficient conditions for a much used formula and implications for the conditional distributions of achieved utility”, Working Paper, Department of Transport Science, KTH Royal Institute of Technology, Stockholm, <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-95397>.
- [9] Lindberg, P.O. (2012b): “Random utility invariance revisited”, Working Paper, Department of Transport Science, KTH Royal Institute of Technology, Stockholm, <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-95400>.
- [10] Lindberg, P.O., E. A. Eriksson, and L.-G. Mattsson (1995): “Invariance of achieved utility in random utility models”, *Environment and Planning A* 27 (1), 121-142.
- [11] Marley, A. (1989): “A random utility family that includes many of the ‘classical’ model and has closed form choice probabilities and reaction times”, *British Journal of Mathematical and Statistical Psychology* 42, 13-36.
- [12] Mattsson, L.-G. and J. Weibull (1981): “Competition and accessibility on a regional labor market”, *Regional Science and Urban Economics* 11, 471-497.
- [13] Mattsson, L.-G., J. Weibull, and P.O. Lindberg (2014): “Extreme values, invariance and choice probabilities”, *Transportation Research Part B* 59, 81-95.
- [14] McFadden, D. (1974): “Conditional logit analysis of qualitative choice behavior”, in Zarembka (ed.), *Frontiers in Econometrics*. New York: Academic Press.
- [15] McFadden, D. (1978): “Modelling the choice of residential location”, Chapter 25 in Karlqvist, Lundqvist, Snickars, Weibull (eds.), *Spatial Interaction Theory and Planning Models*. Amsterdam: North-Holland
- [16] McFadden, D. (1981): “Econometric models of probabilistic choice”, in Manski, McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge (USA): MIT Press.

- [17] McFadden, D. and K. Train (2000): “Mixed MNL models for discrete response”, *Journal of Applied Econometrics* 15, 447-470.
- [18] Nelsen, R. B. (2006): *An Introduction to Copulas, Vol. 2*. Berlin: Springer Series in Statistics.
- [19] Perloff, J. M., and S. C. Salop (1985): “Equilibrium with product differentiation”, *Review of Economic Studies* 52(1), 107–120.
- [20] Resnick, S. and R. Roy (1990a): “Multivariate extremal processes, leader processes and dynamic choice models”, *Advances in Applied Probability* 22 (2), 309-331.
- [21] Resnick, S., and R. Roy (1990b): “Leader and maximum independence for a class of discrete choice models”, *Economics Letters* 33 (3), 259-263.
- [22] Robertson, C. A. and D. J. Strauss (1981): “A characterization theorem for random utility variables”, *Journal of Mathematical Psychology* 23 (2): 184–89.
- [23] Sattinger, M. (1984): “Value of an additional firm in monopolistic competition”, *Review of Economic Studies* 51(2), 321–332.
- [24] Strauss, D. (1979): “Some results on random utility models”, *Journal of Mathematical Psychology* 20 (1): 35–52.
- [25] Train, K., and W. W. Wilson (2008): “Estimation on stated-preference experiments constructed from revealed-preference choices”, *Transportation Research Part B* 42 (3), 191-203.